

# 特許協力条約

PCT

特許性に関する国際予備報告（特許協力条約第二章）

(法第12条、法施行規則第56条)  
〔PCT36条及びPCT規則70〕

出願人又は代理人 の書類記号 P36725-P0	今後の手続きについては、様式PCT/IPEA/416を参照すること。	
国際出願番号 PCT/JP2005/000505	国際出願日 (日.月.年) 17.01.2005	優先日 (日.月.年) 27.01.2004
国際特許分類 (IPC) Int.Cl. G10L13/06(2006.01)		
出願人（氏名又は名称） 松下電器産業株式会社		

1. この報告書は、PCT35条に基づきこの国際予備審査機関で作成された国際予備審査報告である。 法施行規則第57条（PCT36条）の規定に従い送付する。
2. この国際予備審査報告は、この表紙を含めて全部で <u>3</u> ページからなる。
3. この報告には次の附属物件も添付されている。 a. <input checked="" type="checkbox"/> 附属書類は全部で <u>9</u> ページである。  補正されて、この報告の基礎とされた及び／又はこの国際予備審査機関が認めた訂正を含む明細書、請求の範囲及び／又は図面の用紙（PCT規則70.16及び実施細則第607号参照）  第I欄4. 及び補充欄に示したように、出願時における国際出願の開示の範囲を超えた補正を含むものとこの国際予備審査機関が認定した差替え用紙
b. <input type="checkbox"/> 電子媒体は全部で _____ (電子媒体の種類、数を示す)。 配列表に関する補充欄に示すように、電子形式による配列表又は配列表に関連するテーブルを含む。 (実施細則第802号参照)
4. この国際予備審査報告は、次の内容を含む。  <input checked="" type="checkbox"/> 第I欄 国際予備審査報告の基礎 <input type="checkbox"/> 第II欄 優先権 <input type="checkbox"/> 第III欄 新規性、進歩性又は産業上の利用可能性についての国際予備審査報告の不作成 <input type="checkbox"/> 第IV欄 発明の單一性の欠如 <input checked="" type="checkbox"/> 第V欄 PCT35条(2)に規定する新規性、進歩性又は産業上の利用可能性についての見解、それを裏付けるための文献及び説明 <input type="checkbox"/> 第VI欄 ある種の引用文献 <input type="checkbox"/> 第VII欄 国際出願の不備 <input type="checkbox"/> 第VIII欄 国際出願に対する意見

国際予備審査の請求書を受理した日 10.08.2005	国際予備審査報告を作成した日 25.05.2006
名称及びあて先 日本国特許庁 (IPEA/JP) 郵便番号 100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官（権限のある職員） 山下 剛史 5Z 8946 電話番号 03-3581-1101 内線 3541

## 第I欄 報告の基礎

1. 言語に関し、この予備審査報告は以下のものを基礎とした。

- 出願時の言語による国際出願  
 出願時の言語から次の目的のための言語である \_\_\_\_\_ 語に翻訳された、この国際出願の翻訳文  
 国際調査 (PCT規則12.3(a)及び23.1(b))  
 国際公開 (PCT規則12.4(a))  
 国際予備審査 (PCT規則55.2(a)又は55.3(a))

2. この報告は下記の出願書類を基礎とした。(法第6条 (PCT第14条) の規定に基づく命令に応答するために提出された差替え用紙は、この報告において「出願時」とし、この報告に添付していない。)

- 出願時の国際出願書類

- 明細書

第 1, 3-25 ページ、出願時に提出されたもの  
 第 2-2/2 ページ\*、10.08.2005 付けて国際予備審査機関が受理したもの  
 第 \_\_\_\_\_ ページ\*、\_\_\_\_\_ 付けて国際予備審査機関が受理したもの

- 請求の範囲

第 \_\_\_\_\_ 項、出願時に提出されたもの  
 第 \_\_\_\_\_ 項\*、PCT第19条の規定に基づき補正されたもの  
 第 1-7, 11, 18 項\*、10.08.2005 付けて国際予備審査機関が受理したもの  
 第 \_\_\_\_\_ 項\*、\_\_\_\_\_ 付けて国際予備審査機関が受理したもの

- 図面

第 1-16 ページ/図、出願時に提出されたもの  
 第 \_\_\_\_\_ ページ/図\*、\_\_\_\_\_ 付けて国際予備審査機関が受理したもの  
 第 \_\_\_\_\_ ページ/図\*、\_\_\_\_\_ 付けて国際予備審査機関が受理したもの

- 配列表又は関連するテーブル

配列表に関する補充欄を参照すること。

3.  補正により、下記の書類が削除された。

<input type="checkbox"/> 明細書	第 _____	ページ
<input checked="" type="checkbox"/> 請求の範囲	第 <u>8-10, 12-17</u>	項
<input type="checkbox"/> 図面	第 _____	ページ/図
<input type="checkbox"/> 配列表 (具体的に記載すること)	_____	
<input type="checkbox"/> 配列表に関連するテーブル (具体的に記載すること)	_____	

4.  この報告は、補充欄に示したように、この報告に添付されかつ以下に示した補正が出願時における開示の範囲を超えてされたものと認められるので、その補正がされなかったものとして作成した。(PCT規則70.2(c))

<input type="checkbox"/> 明細書	第 _____	ページ
<input type="checkbox"/> 請求の範囲	第 _____	項
<input type="checkbox"/> 図面	第 _____	ページ/図
<input type="checkbox"/> 配列表 (具体的に記載すること)	_____	
<input type="checkbox"/> 配列表に関連するテーブル (具体的に記載すること)	_____	

\* 4. に該当する場合、その用紙に "superseded" と記入されることがある。

第V欄 新規性、進歩性又は産業上の利用可能性についての法第12条（PCT35条(2)）に定める見解、それを裏付ける文献及び説明

1. 見解

新規性 (N)	請求の範囲 <u>1 - 7, 11, 18</u>	有
	請求の範囲 _____	無
進歩性 (I S)	請求の範囲 <u>1 - 7, 11, 18</u>	有
	請求の範囲 _____	無
産業上の利用可能性 (I A)	請求の範囲 <u>1 - 7, 11, 18</u>	有
	請求の範囲 _____	無

2. 文献及び説明 (PCT規則70.7)

文献1 : JP 9-244693 A (エヌ・ティ・ティ・データ通信株式会社)  
1997.09.19, 全文, 図1-10 (ファミリーなし)

文献2 : JP 2002-351489 A (株式会社ナムコ)  
2002.12.06, 全文, 図1-24 (ファミリーなし)

文献3 : JP 4-158397 A (株式会社エイ・ティ・アール自動翻訳電話研究所)  
1992.06.01, 全文, 第1-3図 (ファミリーなし)

請求の範囲1-7, 11, 18に係る発明におけるモーフィング手段（ステップ）の具体的な構成（記憶されている各音声素片情報の声質を示す固定点をN次元の座標上に配置して表示するとともに、ユーザの操作に基づいて複数の設定点を前記座標上に配置して表示し、前記複数の設定点間を時系列に沿って連続的に移動する移動点および前記固定点の配置に基づいて、モーフィングに寄与する複数の合成音声情報のそれぞれの、時系列に沿って変化する割合を導出して指定し、生成された複数の合成音声情報のそれぞれを、指定された時系列に沿って変化する割合だけ用いることにより、テキストデータに含まれる文字に対応した、複数の声質の中間的な声質の合成音声を示す中間合成音声信号を生成する点）については、上記文献1-3等の国際調査報告において引用されたいずれの文献にも記載も示唆もされておらず、該構成は当業者と言えども容易に想到し得るものではない。

ピーク)を特定して、その部位を基準にモーフィング処理を行うが、その部位を誤って特定してしまうことがある。その結果、生成された合成音声の音質が悪くなってしまう。そこで、本発明は、このような問題に鑑みてなされたものであって、声質の自由度が広く良い音質の合成音声をテキストデータから生成する音声合成装置を提供することを目的とする。

### 課題を解決するための手段

- [0010] 上記目的を達成するために、本発明に係る音声合成装置は、互いに異なる声質ごとに、当該声質に属する複数の音声素片に関する音声素片情報を予め記憶している記憶手段と、テキストデータを取得するとともに、前記記憶手段に記憶されている複数の音声素片情報から、前記声質ごとに、前記テキストデータに含まれる文字に対応した当該声質の合成音声を示す合成音声情報を生成する音声情報生成手段と、前記記憶手段に記憶されている各音声素片情報の声質を示す固定点をN次元(Nは自然数)の座標上に配置して表示するとともに、ユーザの操作に基づいて複数の設定点を前記座標上に配置して表示し、前記複数の設定点間を時系列に沿って連続的に移動する移動点および前記固定点の配置に基づいて、モーフィングに寄与する前記複数の合成音声情報のそれぞれの、時系列に沿って変化する割合を導出して指定する指定手段と、前記音声情報生成手段により生成された複数の合成音声情報のそれぞれを、前記指定手段により指定された時系列に沿って変化する割合だけ用いることにより、前記テキストデータに含まれる文字に対応した、前記複数の声質の中間的な声質の合成音声を示す中間合成音声情報を生成するモーフィング手段と、前記モーフィング手段によって生成された前記中間合成音声情報を前記中間的な声質の合成音声に変換して出力する音声出力手段とを備え、前記音声情報生成手段は、前記複数の合成音声情報をそれぞれ複数の特徴パラメタの列として生成し、前記モーフィング手段は、前記複数の合成音声情報の互いに対応する特徴パラメタの中間値を計算することで、前記中間合成音声情報を生成することを特徴とする。これにより、例えば、第1の声質に対する第1の音声素片情報、及び第2の声質に対する第2の音声素片情報だけを記憶手段に予め記憶させておけば、第1及び第2の声質の中間的な声質の合成音声が出力されるため、記憶手段に予め記憶させて

おく内容の声質に限定されずに声質の自由度を広めることができる。また、第1及び第2の声質を有する第1及び第2の合成音声情報を基礎に中間合成音声情報が生成されるため、従来例のようにスペクトルのダイナミックレンジを大きくしすぎるような処理がなされず、合成音声の音質を良い状態に維持することができる。また、本発明に係る音声合成装置は、テキストデータを取得して、そこに含まれる文字列に応じた合成音声を出力するため、ユーザに対する使い勝手を向上することができる。さらに、本発明に係る音声合成装置は、第1及び第2の合成音声情報の互いに対応する特徴パラメタの中間値を計算して中間合成音声情報を生成するため、従来例のように2つのスペクトルをモーフィング処理する場合と比べて、基準とする部位を誤って特定してしまうことなく、合成音声の音質を良くすることができ、さらに、計算量を軽減することができる。さらに、本発明に係る音声合成装置は、固定点とユーザの操作に基づいて配置される設定点とに従って複数の合成音声情報のモーフィングに寄与する割合が変化するため、ユーザは音声素片情報の声質に対する類似度を容易に入力することができる。

また、本発明に係る音声合成装置は、第1の声質に属する複数の音声素片に関する第1の音声素片情報、及び前記第1の声質と異なる第2の声質に属する複数の音声素片に関する第2の音声素片情報を予め記憶している記憶手段と、テキストデータを取得するとともに、前記記憶手段の第1の音声素片情報から、前記テキストデータに含まれる文字に対応した前記第1の声質の合成音声を示す第1の合成音声情報を生成し、前記記憶手段の第2の音声素片情報から、前記テキストデータに含まれる文字に対応した前記第2の声質の合成音声を示す第2の合成音声情報を生成する音声情報生成手段と、前記音声情報生成手段により生成された前記第1及び第2の合成音声情報から、前記テキストデータに含まれる文字に対応した、前記第1及び第2の声質の中間的な声質の合成音声を示す中間合成音声情報を生成するモーフィング手段と、前記モーフィング手段によって生成された前記中間合成音声情報を前記中間的な声質の合成音声に変換して出力する音声出力手段とを備え、前記音声情報生成手段は、前記第1及び第2の合成音声情報をそれぞれ複数の特徴パラメタの列として生成し、前記モーフィング手段は、前記第1及び第2の合成音声情報の互

いに対応する特徴パラメタの中間値を計算することで、前記中間合成音声情報を生成することを特徴とする。

[0011] これにより、第1の声質に対する第1の音声素片情報、及び第2の声質に対する第2の音声素片情報だけを記憶手段に予め記憶させておけば、第1及び第2の声質の中間的な声質の合成音声が出力されるため、記憶手段に予め記憶させておく内容の声質に限定されずに声質の自由度を広めることができる。また、第1及び第2の声質を有する第1及び第2の合成音声情報を基礎に中間合成音声情報が生成されるため、従来例のようにスペクトルのダイナミックレンジを大きくしそぎるような処理がなさ

## 請求の範囲

- [1] (補正後) 互いに異なる声質ごとに、当該声質に属する複数の音声素片に関する音声素片情報を予め記憶している記憶手段と、  
テキストデータを取得するとともに、前記記憶手段に記憶されている複数の音声素片情報から、前記声質ごとに、前記テキストデータに含まれる文字に対応した当該声質の合成音声を示す合成音声情報を生成する音声情報生成手段と、  
前記記憶手段に記憶されている各音声素片情報の声質を示す固定点をN次元(Nは自然数)の座標上に配置して表示するとともに、ユーザの操作に基づいて複数の設定点を前記座標上に配置して表示し、前記複数の設定点間を時系列に沿って連続的に移動する移動点および前記固定点の配置に基づいて、モーフィングに寄与する前記複数の合成音声情報のそれぞれの、時系列に沿って変化する割合を導出して指定する指定手段と、  
前記音声情報生成手段により生成された複数の合成音声情報を、前記指定手段により指定された時系列に沿って変化する割合だけ用いることにより、前記テキストデータに含まれる文字に対応した、前記複数の声質の中間的な声質の合成音声を示す中間合成音声情報を生成するモーフィング手段と、  
前記モーフィング手段によって生成された前記中間合成音声情報を前記中間的な声質の合成音声に変換して出力する音声出力手段と  
を備え、  
前記音声情報生成手段は、前記複数の合成音声情報をそれぞれ複数の特徴パラメタの列として生成し、  
前記モーフィング手段は、前記複数の合成音声情報の互いに対応する特徴パラメタの中間値を計算することで、前記中間合成音声情報を生成することを特徴とする音声合成装置。
- [2] (補正後) 前記モーフィング手段は、前記音声出力手段から出力される合成音声の声質がその出力中に連続的に変化するように、前記複数の合成音声情報の前記中間合成音声情報に対して寄与する割合を変化させる  
ことを特徴とする請求項1記載の音声合成装置。

- [3] (補正後)前記記憶手段は、前記音声素片情報ごとに、当該音声素片情報により示される各音声素片における基準を示す内容の特徴情報を、当該音声素片情報に含めて記憶しており、  
前記音声情報生成手段は、前記複数の合成音声情報を、それぞれに前記特徴情報を含めて生成し、  
前記モーフィング手段は、前記複数の合成音声情報を、それぞれに含まれる前記特徴情報によって示される基準を用いて整合した上で前記中間合成音声情報を生成する  
ことを特徴とする請求項1記載の音声合成装置。
- [4] (補正後)前記基準は、前記複数の音声素片情報のそれぞれにより示される各音声素片の音響的特徴の変化点である  
ことを特徴とする請求項3記載の音声合成装置。
- [5] (補正後)前記音響的特徴の変化点は、前記複数の音声素片情報のそれぞれに示される各音声素片をHMM(Hidden Markov Model)で表した最尤経路上の状態遷移点であつて、  
前記モーフィング手段は、前記複数の合成音声情報を、前記状態遷移点を用いて時間軸上で整合した上で前記中間合成音声情報を生成する  
ことを特徴とする請求項4記載の音声合成装置。
- [6] (補正後)前記音声合成装置は、さらに、  
前記声質ごとに、当該声質に対応する画像を示す画像情報を予め記憶している画像記憶手段と、  
前記複数の画像情報のそれぞれにより示される画像の中間的な画像であつて、前記中間合成音声情報の声質に対応する画像を示す中間画像情報を、前記複数の画像情報から生成する画像モーフィング手段と、  
前記画像モーフィング手段により生成された中間画像情報を取得して、前記中間画像情報により示される画像を、前記音声出力手段から出力される合成音声に同期させて表示する表示手段と  
を備えることを特徴とする請求項1記載の音声合成装置。

[7] (補正後)前記複数の画像情報はそれぞれ前記声質に対応する顔画像を示すことを特徴とする請求項6記載の音声合成装置。

[8] (削除)

[9] (削除)

[10] (削除)

[11] (補正後)互いに異なる声質ごとに、当該声質に属する複数の音声素片に関する音声素片情報を予め記憶しているメモリを用いることで、合成音声を生成して出力する音声合成方法であって、

テキストデータを取得するテキスト取得ステップと、

前記メモリの複数の音声素片情報から、前記声質ごとに、前記テキストデータに含まれる文字に対応した当該声質の合成音声を示す合成音声情報を生成する音声情報生成ステップと、

前記メモリに記憶されている各音声素片情報の声質を示す固定点をN次元(Nは自然数)の座標上に配置して表示するとともに、ユーザの操作に基づいて複数の設定点を前記座標上に配置して表示し、前記複数の設定点間を時系列に沿って連続的に移動する移動点および前記固定点の配置に基づいて、モーフィングに寄与する前記複数の合成音声情報のそれぞれの、時系列に沿って変化する割合を導出して指定する指定ステップと、

前記音声情報生成ステップで生成された複数の合成音声情報のそれぞれを、前記

指定ステップで指定された時系列に沿って変化する割合だけ用いることにより、前記テキストデータに含まれる文字に対応した、前記複数の声質の中間的な声質の合成音声を示す中間合成音声情報を生成するモーフィングステップと、

前記モーフィングステップで生成された前記中間合成音声情報を前記中間的な声質の合成音声に変換して出力する音声出力ステップと  
を含み、

前記音声情報生成ステップでは、前記複数の合成音声情報をそれぞれ複数の特徴パラメタの列として生成し、

前記モーフィングステップでは、前記複数の合成音声情報の互いに対応する特徴パラメタの中間値を計算することで、前記中間合成音声情報を生成することを特徴とする音声合成方法。

[12] (削除)

[13] (削除)

[14] (削除)

[15] (削除)

[16] (削除)

[17] (削除)

- [18] (補正後) 互いに異なる声質ごとに、当該声質に属する複数の音声素片に関する音声素片情報を予め記憶しているメモリを用いて、合成音声を生成して出力するためのプログラムであって、  
テキストデータを取得するテキスト取得ステップと、  
前記メモリの複数の音声素片情報から、前記声質ごとに、前記テキストデータに含まれる文字に対応した当該声質の合成音声を示す合成音声情報を生成する音声情報生成ステップと、  
前記メモリに記憶されている各音声素片情報の声質を示す固定点をN次元(Nは自然数)の座標上に配置して表示するとともに、ユーザの操作に基づいて複数の設定点を前記座標上に配置して表示し、前記複数の設定点間を時系列に沿って連続的に移動する移動点および前記固定点の配置に基づいて、モーフィングに寄与する前記複数の合成音声情報をそれぞれの、時系列に沿って変化する割合を導出して指定する指定ステップと、  
前記音声情報生成ステップで生成された複数の合成音声情報を、前記指定ステップで指定された時系列に沿って変化する割合だけ用いることにより、前記テキストデータに含まれる文字に対応した、前記複数の声質の中間的な声質の合成音声を示す中間合成音声情報を生成するモーフィングステップと、  
前記モーフィングステップで生成された前記中間合成音声情報を前記中間的な声質の合成音声に変換して出力する音声出力ステップと  
をコンピュータに実行させ、  
前記音声情報生成ステップでは、前記複数の合成音声情報をそれぞれ複数の特

微パラメタの列として生成し、

前記モーフィングステップでは、前記複数の合成音声情報の互いに対応する特徴パラメタの中間値を計算することで、前記中間合成音声情報を生成することを特徴とするプログラム。